|  |  |
| --- | --- |
| Activity | Data Type |
| Number of beatings from Wife | Discrete |
| Results of rolling a dice | Ordinal |
| Weight of a person | Ratio |
| Weight of Gold | Ratio |
| Distance between two places | Ratio |
| Length of a leaf | Ratio |
| Dog's weight | Ratio |
| Blue Color | Nominal |
| Number of kids | Discrete |
| Number of tickets in Indian railways | Ordinal |
| Number of times married | Discrete |
| Gender (Male or Female) | Nominal |

**Q1) Identify the Data type for the Following:**

**Q2) Identify the Data types, which were among the following**

Nominal, Ordinal, Interval, Ratio.

|  |  |
| --- | --- |
| Data | Data Type |
| Gender | Nominal |
| High School Class Ranking | Ordinal |
| Celsius Temperature | Interval |
| Weight | Ratio |
| Hair Color | Nominal |
| Socioeconomic Status | Ordinal |
| Fahrenheit Temperature | Ordinal |
| Height | Ratio |
| Type of living accommodation | Ordinal |
| Level of Agreement | Ordinal |
| IQ(Intelligence Scale) | Ratio |
| Sales Figures | Interval |
| Blood Group | Nominal |
| Time Of Day | Ratio |
| Time on a Clock with Hands | Ratio |
| sNumber of Children | Ordinal |
| Religious Preference | Nominal |
| Barometer Pressure | Ratio |
| SAT Scores | Ratio |
| Years of Education | Interval |

**Q3) Three Coins are tossed, find the probability that two heads and one tail are obtained?**

Ans: **PROBABILITY HAPPENS**

**HHH, HHT, HTT, TTT, TTH, THH, HTH, THT**

**No of possibility obtain = 8**

**No. of events that 2 Head & 1 tail are obtained = 3**

**Probability = 3/8**

**Q4) Two Dice are rolled, find the probability that sum is**

1. **Equal to 1**
2. **Less than or equal to 4**
3. **Sum is divisible by 2 and 3**
4. **Ans.**

(1,1) (1,2) (1,3) (1,4) (1,5) (1,6)

(2,1) (2,2) (2,3) (2,4) (2,5) (2,6)

(3,1) (3,2) (3,3) (3,4) (3,5) (3,6)

(4,1) (4,2) (4,3) (4,4) (4,5) (4,6)

(5,1) (5,2) (5,3) (5,4) (5,5) (5,6)

(6,1) (6,2) (6,3) (6,4) (6,5) (6,6)

1. Equal to 1 = Probability 0
2. Less than or equal to 4 = Probability 6/36 = 1/6
3. Sum is divisible by 2 and 3 = Probability 23/36

Sum is divisible by both 2 and 3 together in same pair = Probability 6/36=1/6

**Q5) A bag contains 2 red, 3 green and 2 blue balls. Two balls are drawn at random. What is the probability that none of the balls drawn is blue?**

Total number of balls  
= (2 + 3 + 2)  
= 7  
Let S be the sample space  
Then, n(S) = Number of ways of drawing 2 balls out of 7

n(S)=7C2n

(S)=(7×6)(2×1)

n(S)=21

n(S)=7C2

n(S)=(7×6)(2×1)

n(S)=21

Let E = Event of 2 balls, none of which is blue  
∴ n(E) = Number of ways of drawing 2 balls out of (2 + 3) balls

n(E)=5C2

n(E)=(5×4)/(2×1)

n(E)=10

∴P(E)=n(E)/n(S)=10/21

**Q6) Calculate the Expected number of candies for a randomly selected child**

**Below are the probabilities of count of candies for children (ignoring the nature of the child-Generalized view)**

|  |  |  |
| --- | --- | --- |
| CHILD | Candies count | Probability |
| A | 1 | 0.015 |
| B | 4 | 0.20 |
| C | 3 | 0.65 |
| D | 5 | 0.005 |
| E | 6 | 0.01 |
| F | 2 | 0.120 |

Child A – probability of having 1 candy = 0.015.

Child B – probability of having 4 candies = 0.20

Ans: Expected number of candies for a randomly selected child  = **3.09**

Step-by-step explanation:

Expected number of candies for a randomly selected child

=  (1 \* 0.015)  + ( 4\*0.20 ) + (3 \*0.65)  + (5\*0.005 ) +( 6 \*0.01 ) + (2 \* 0.12)

= 0.015 + 0.8  + 1.95 + 0.025 + 0.06 + 0.24

=   3.090

=  3.09

**Q7) Calculate Mean, Median, Mode, Variance, Standard Deviation, Range & comment about the values / draw inferences, for the given dataset**

* **For Points,Score,Weigh>**

**Find Mean, Median, Mode, Variance, Standard Deviation, and Range and also Comment about the values/ Draw some inferences.**

**Use Q7.csv file**

**ANSWER:**

**POINTS**

Mean: 3.596563

Median: 3.695

Mode: 3.92

Variance: 0.2858814

Standard Deviation: 0.5346787

Range: 2.76 4.93

**SCORE:**

Mean: 3.21725

Median: 3.325

Mode: 3.44

Variance: 0.957379

Standard Deviation: 0.9784574

Range: 1.513 5.424

**WEIGH:**

Mean: 17.84875

Median: 17.71

Mode: 17.02

Variance: 3.193166

Standard Deviation: 1.786943

Range: 14.5 22.9

**Q8) Calculate Expected Value for the problem below**

1. **The weights (X) of patients at a clinic (in pounds), are**

**108, 110, 123, 134, 135, 145, 167, 187, 199**

**Assume one of the patients is chosen at random. What is the Expected Value of the Weight of that patient?**

Ans. Expected value of Weight=(108+110+123+134+135+145+167+187+199)/9

**=145.33**

**Q9) Calculate Skewness, Kurtosis & draw inferences on the following data**

**Cars speed and distance**

**Use Q9\_a.csv**

|  |  |  |  |
| --- | --- | --- | --- |
|  | **INDEX** | **SKEWNESS** | **KURTOSIS** |
| **SPEED** | **0.000** | **-0.1139548** | **2.422853** |
| **DISTANCE** | **1.799040** | **0.78242835** | **3.248019** |

**SP and Weight(WT)**

**Use Q9\_b.csv**

|  |  |  |  |
| --- | --- | --- | --- |
|  | **INDEX** | **SKEWNESS** | **KURTOSIS** |
| **SP** | **0.000** | **1.5814537** | **5.723521** |
| **WEIGHT** | **1.799634** | **-0.6033099** | **3.819466** |

**Q10) Draw inferences about the following boxplot & histogram**

ANSWER :

Histogram: We can see positive skewness here. Thus mass of distribution is concentrated on the left.

Boxplot: We can see lots of outliers or error values towards upper extreme.



**Q11)** Suppose we want to estimate the average weight of an adult male in Mexico. We draw a random sample of 2,000 men from a population of 3,000,000 men and weigh them. We find that the average person in our sample weighs 200 pounds, and the standard deviation of the sample is 30 pounds. Calculate 94%,98%,96% confidence interval?

Confidence interval=interval estimation of parameter

=point estimation+-margin of error

=X bar+-Z1-a\*sigma/root n

=200+-Z94\*30/root2000

=200+-qnorm(0.97)\*30/root2000

=200+-1.88\*0.67

Positive value=201.26, negative value=198.74 For 98% Confidence interval,

Positive value=201.56, Negative value=198.44 For 96% Confidence interval,

Positive value=201.38, Negative value= 198.62

**Q12)** Below are the scores obtained by a student in tests

**34,36,36,38,38,39,39,40,40,41,41,41,41,42,42,45,49,56**

1. Find mean, median, variance, standard deviation.

**Ans:**

**Mean: 41**

**Median: 40.5**

**Variance: 25.52941**

**Standard Deviation: 5.052664**

1. What can we say about the student marks?

Q13) What is the nature of skewness when mean, median of data are equal?

**Ans.** When mean and median is equal then we find no Skewness.

Q14) What is the nature of skewness when mean > median ?

**Ans.** When mean is greater than median then we find right skewness.

Q15) What is the nature of skewness when median > mean?

Ans. When median is greater than mean then we find left skewness.

Q16) What does positive kurtosis value indicates for a data ?

Ans. More ‘Peakedness’ of data.

Q17) What does negative kurtosis value indicates for a data?

Ans. Less ‘peakedness’ of data.

Q18) Answer the below questions using the below boxplot visualization.



What can we say about the distribution of the data?

What is nature of skewness of the data?

What will be the IQR of the data (approximately)?   
  
Ans.

* + Area under the entire curve is not equals to 1 thus it is not normal distribution of data.
  + Tails of the data starts from the left side of the graph and we are getting IQ part towards the right side of the graph so the skewness type is left skewed.
  + IQR=Upper Quartile-Lower Quartile=(18-10)=8

Q19) Comment on the below Boxplot visualizations?



Draw an Inference from the distribution of data for Boxplot 1 with respect Boxplot 2.

**Ans.** Two box plot have same median value approx. 262. IQR of first boxplot= 25

IQR of second boxplot= 75

Distribution type= normal distribution as no left skewed or right skewed plot has been found.

There are no outliers found.

Q 20) Calculate probability from the given dataset for the below cases

Data \_set: Cars.csv

Calculate the probability of MPG of Cars for the below cases.

MPG <- Cars$MPG

* 1. P(MPG>38)
  2. P(MPG<40)
  3. P (20<MPG<50)

Ans.

Mean= 34.42208

Standard deviation= 9.131445

When , MPG > 38

PNorm = 0.6251

Probability = (1 – 0.65 )

= 0.35

When , MPG < 40,

PNorm = 0.7294571

When, (20 < MPG < 50)

PNorm = 0.8989178

Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

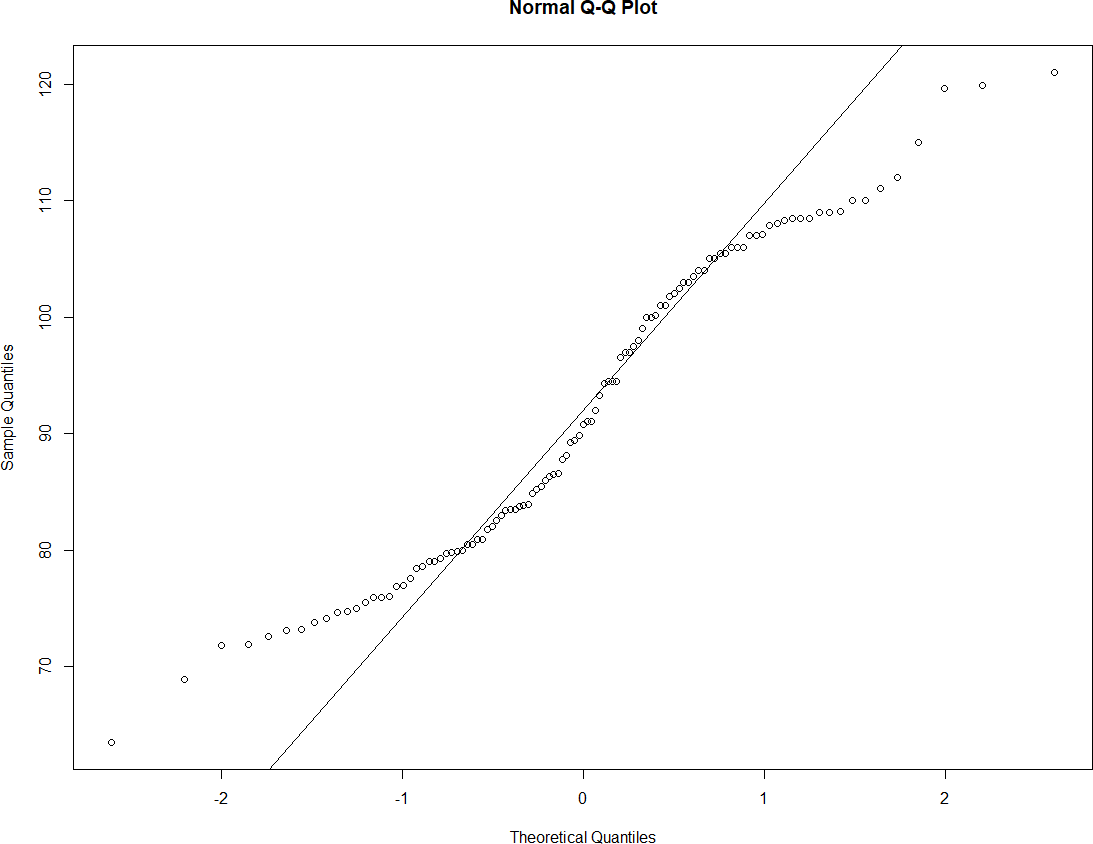
Dataset: Cars.csv

Ans. From QQ plot we can see normal distribution of data.

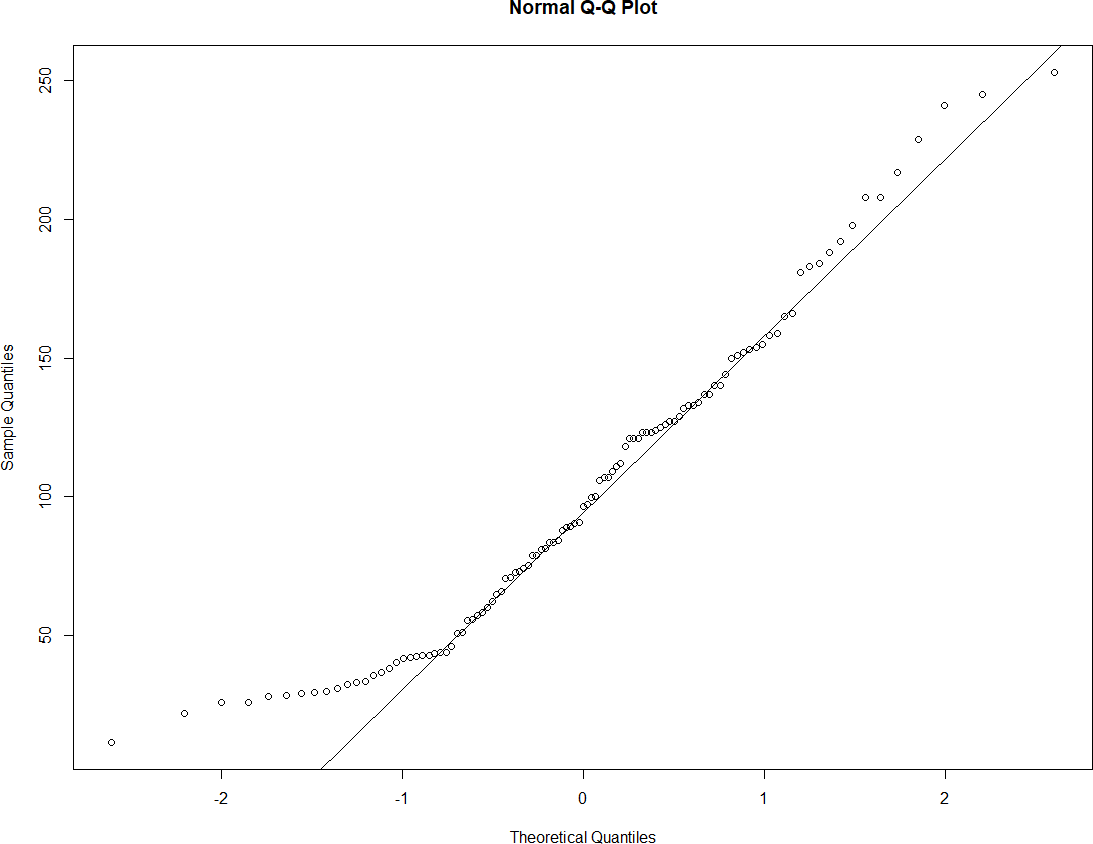
1. Check Whether the Adipose Tissue (AT) and Waist Circumference(Waist) from wc-at data set follows Normal Distribution

Dataset: wc-at.csv

Ans. Waist: From QQ plot we can see normal distribution of data.



Adipose Tissue (AT): From QQ plot we can see normal distribution of data.



Q 22) Calculate the Z scores of 90% confidence interval,94% confidence interval, 60% confidence interval

Ans. Z scores of 90% confidence interval,

Positive value= 1.644854, Negative value= 1.036433

94% confidence interval,

Positive value= 1.880794, Negative value= 1.340755

60% confidence interval,

Positive value= 0.8416212, Negative value= -0.2533471

Q 23) Calculate the t scores of 95% confidence interval, 96% confidence interval, 99% confidence interval for sample size of 25

Ans.

95% confidence interval= 2.063899

96% confidence interval= 2.171545

99% confidence interval= 2.79694

Q 24**)** A Government company claims that an average light bulb lasts 270 days. A researcher randomly selects 18 bulbs for testing. The sampled bulbs last an average of 260 days, with a standard deviation of 90 days. If the CEO's claim were true, what is the probability that 18 randomly selected bulbs would have an average life of no more than 260 days

Hint:

rcode 🡪 pt(tscore,df)

df 🡪 degrees of freedom

Ans. X bar=260, SD=90, n=18, MU=270

Tscore= (260-270)/90/root 18 =-0.47

pt(-0.47,17)= 0.3221639

Probability 32%